Chapter 5

Effectiveness of screening in populations

This chapter deals with population measures of effectiveness of cervical cancer screening. At the population level, effectiveness is ultimately assessed by the reduction in mortality due to cancer of the cervix. There are two reasons to expect mortality to decrease as a result of screening: removal of incident cases through detection and treatment of premalignant lesions, and diagnosis of invasive lesions at earlier, more curable stages. Because screening for cervical cancer can detect precursor lesions that can be treated to prevent progression to invasive disease, reduction of cervical cancer incidence can be used as a measure of effectiveness as well.

Four methods have been used to assess the effectiveness of screening: individual-based studies using case–control or cohort designs (see Chapter 4); ecological analyses (correlating screening activity with changes in mortality or incidence rates across time, place or age group); modelling of screening policy and practice to estimate effectiveness; and evaluation of operational parameters of screening. The latter includes screening performance indicators such as participation, quality and adequacy of follow-up of positive test results. This chapter is concerned with the last three of these.

Incidence and mortality trends in relation to screening

Time trends are of considerable interest, in part for the light that may be shed on changes in exposure to etiological factors (especially between women of different generations) and in part as a means of evaluating the success, or otherwise, of screening programmes. Because of their comprehensive coverage and availability, mortality data are often used in studies of time trends; however, care is needed in doing so, on account of the changing proportions of deaths assigned to 'Uterus, unspecified' (NOS) (see Chapter 1), and possible changes in treatment-induced survival, which may be quite large if long time series are studied (Pontén et al., 1995).

A reduction over time in the incidence of invasive cervical cancer, especially in those age groups where screening is mostly targeted, is another long-term indicator of effectiveness. However, high-quality population-based incidence data, as provided by population-based cancer registries, are available in relatively few regions of the world (Parkin et al., 2002) and fewer still have incidence data covering extended periods of time.

Time trends by region

Until quite recently, most studies focused on the overall cervical cancer trends rather than looking separately at adenocarcinoma and squamous-cell cancer. However, cytological screening identifies mainly the latter. Since most cervical cancers are squamous-cell carcinomas, studies of overall cancer incidence and mortality largely reflect trends in this histological type.

Developed countries
Europe
Trends in cervical cancer incidence and mortality have been intensively studied in the Nordic countries, where it has also been possible to compare the trends in the different countries in relation to the intensity of screening undertaken (Hakama, 1982; Hakulinen et al., 1986, Läärä et al., 1987; Engeland et al., 1993; Sigurdsson, 1993, 1995; Hristova & Hakama, 1997; Anttila & Läärä, 2000; Moller et al., 2002). In these countries, national incidence and mortality data are available from before and after the times that screening programmes were implemented. Towards the end of the 1960s, Finland, Sweden and Iceland had nationwide, organized screening programmes, and the same was true for several Danish counties. Norway, in contrast, had organized screening only in a single county covering about 5% of the population. Throughout the Nordic countries, opportunistic testing also increased at the same time.

From the late 1960s, a decrease was seen in both the incidence of and mortality from cervical cancer in Finland, Sweden, Iceland and Denmark (Figure 54). The decrease, relative to the levels before screening, was largest in Finland, where the age-standardized mortality rate decreased more than 80% from 6.6 deaths per 100 000 in early 1960s to 1.2 deaths per 100 000 in the early 1990s (decrease 82%) (rates adjusted for age to the world standard population). In the earlier period, women 30–55 years of age were invited with a five-year screening interval and it was only in the early 1990s that the maximum age for invitation was raised to 60 years in Finland. The decreases in the mortality rates were 65% and 55%, respectively, in Sweden and Denmark, with partial coverage by organized programmes. A reduction in cervical cancer incidence was also observed in the Danish counties with organized screening compared to those without (Lynge *et al.*, 1989). In Norway, the incidence

increased until the mid-1970s, and the decrease in mortality was considerably less (41% from the early 1960s to the early 1990s) than in the other Nordic countries. At that time, opportunistic screening had become frequent also in Norway. A national organized programme of cervical cancer screening started in Norway in 1995 (Nygard *et al.*, 2002). The trend in incidence was quite similar to the trend in mortality within each country up to the mid-1990s in terms of percentage reduction in the age-standardized rate. Also the incidence to mortality ratios were quite stable.

In general, incidence and mortality have also declined in the last 20–40 years in many other European countries (Coleman *et al.*, 1993; Beral *et al.*, 1994), but in some populations increases have been observed among younger women (aged under 35 years), particularly during the 1970s and 1980s (Figure 55). This was first noted in England and Wales, where generations of women born since

about 1935 were observed to be at increasingly high risk (Hill & Adelstein, 1967; Cook & Draper, 1984; Parkin *et al.*, 1985). Similar phenomena have been seen in Belgium (Vyslouzilova *et al.*, 1997), Slovenia (Kirn *et al.*, 1992), Slovakia (Vlasak *et al.*, 1991), Spain (Llorca *et al.*, 1999) and in several other countries of eastern Europe (Beral *et al.*, 1994).

Since the early 1990s, the incidence rate has started to increase in Finland among women below 55 years of age (Figure 56) (Anttila *et al.*, 1999). This trend is probably due to a combination of changes in sexual lifestyles and increased transmission of papillomaviruses in younger generations of women, as well as inadequacies in the screening programme such as changes in laboratory procedures during this time (Nieminen *et al.*, 2002). Because the effect of increasing incidence has been partly obscured by the protective effect of screening, in some countries, there has been little or no increase in
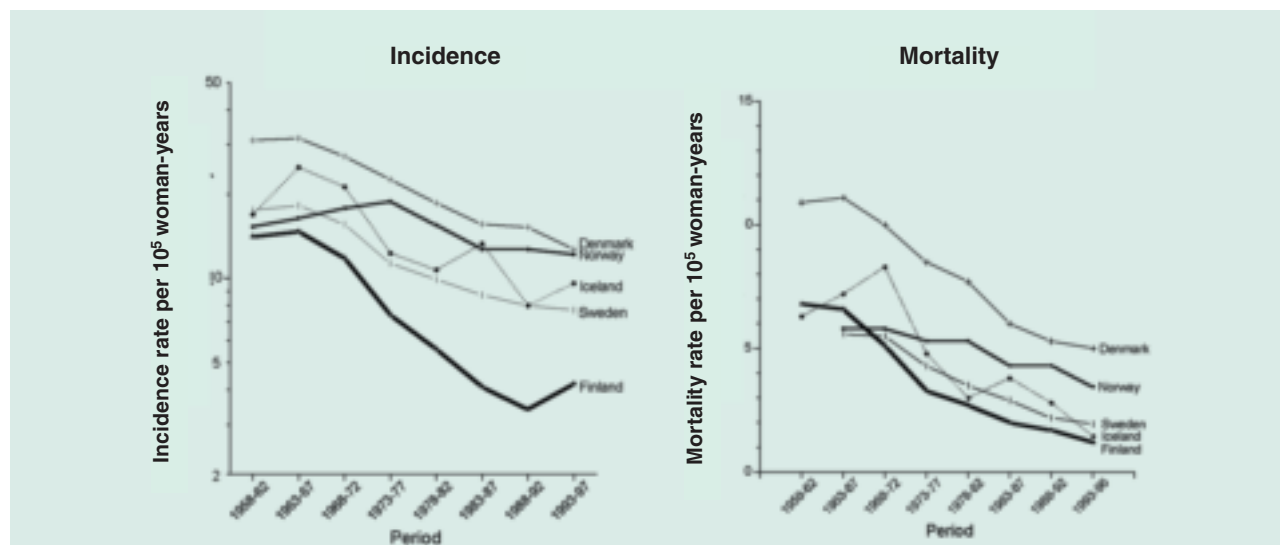

Figure 54 Incidence and mortality rates of cervical cancer in the Nordic countries, 1958–97 (mortality available up to 1996)
Whole female population, adjusted for age to the world standard population (Läärä *et al.,* 1987; Engeland *et al.*, 1993; Hristova & Hakama 1997; Parkin *et al.*, 1997; Moller *et al.,* 2002; EUROCIM (European Network of Cancer Registries) database).
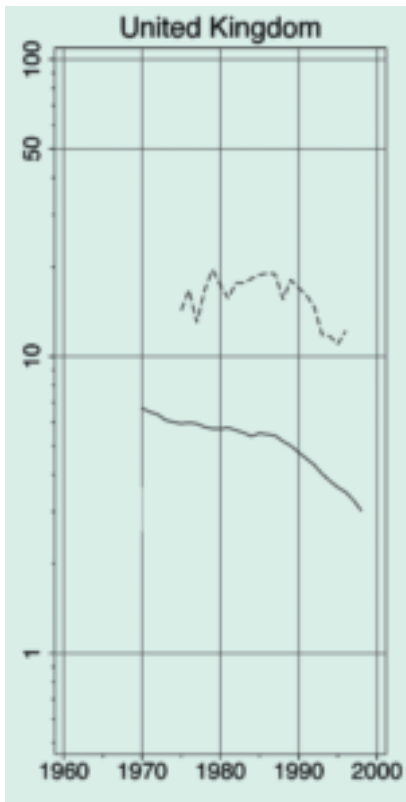
United Kingdom

Figure 55 Cervical cancer incidence (---) and mortality (—) trends in the United Kingdom, all ages
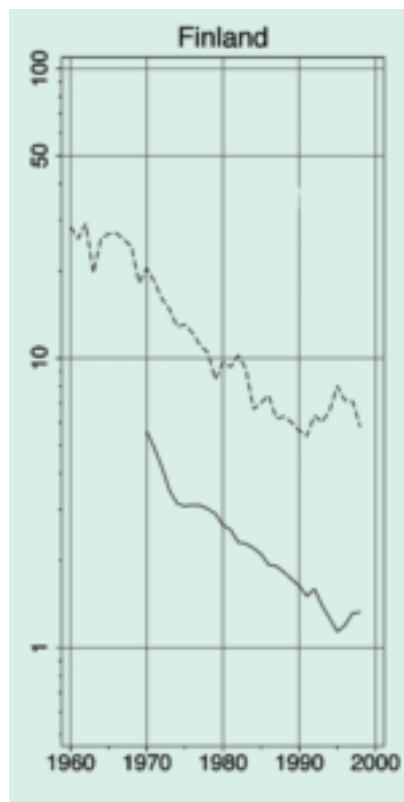
Finland

Figure 56 Cervical cancer incidence (----) and mortality (—) trends in Finland, all ages
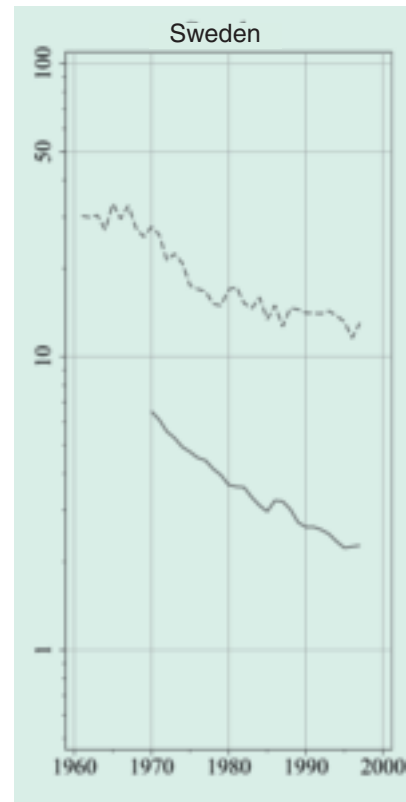
Sweden

Figure 57 Cervical cancer incidence (----) and mortality (—) trends in Sweden, all ages
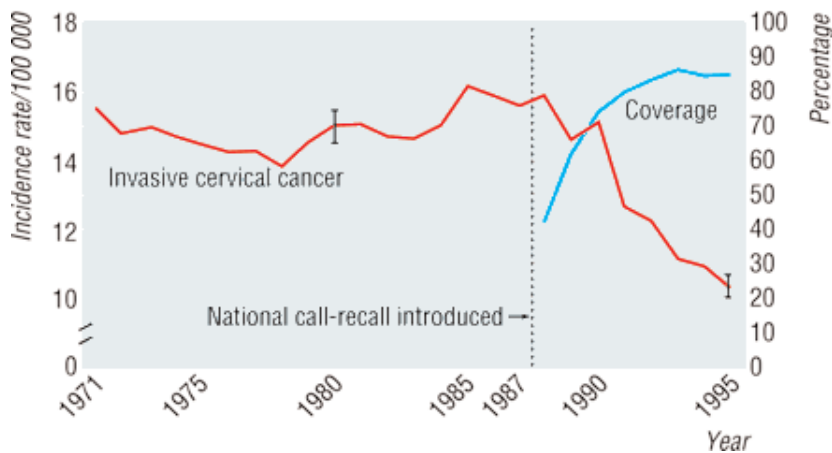
Figure 58 Age-standardized incidence of invasive cervical cancer, Englandl, 1971–95

From Quinn *et al.* (1999) (reproduced with permission from the BMJ Publishing Group).

risk in young women (for example, Sweden (Figure 57); Bergström *et al.*, 1999).

In the United Kingdom, cytological screening was introduced in the 1960s, but an organized programme, including a call/recall system and quality assurance, was implemented only from the 1988 onwards, leading to increased coverage within the targeted population. A sharp decrease in cervical cancer incidence and mortality rates since 1990 has been attributed to this organized programme (Sasieni *et al.*, 1995, Gibson *et al.*, 1997; Quinn *et al.*, 1999; Sasieni & Adams, 1999) (Figure 58). The average drop in the age-adjusted mortality rate was estimated as 1–2% per year during

1960–88 and 7% since then (Sasieni *et al.*, 1995).

Coding of deaths as due to cancer of the uterus NOS has been common in many countries and this affects comparability over time. In Belgium, an attempt has been made to estimate the proportions of deaths ascribed to cancer of the uterus NOS that should be redistributed to cervix and other uterine cancer (Arbyn & Geys, 2002). The corrected age-standardized mortality rates decreased from 14 per 100 000 in the 1950s to 4.5 in the 1990s (68% decrease), while the certified rates decreased from 6.3 to 3 (52% decrease).

In a number of eastern European countries such as Bulgaria, Romania and the Russian Federation, where little or no screening has taken place, cervical cancer mortality rates are rapidly rising (Figure 59), notably among recently born generations, as seen for Bulgarian women. In more affluent eastern European countries such as the Czech Republic, Hungary and Poland, there is some evidence of very recent declines (Figure 59), and in terms of birth cohort, mortality may have peaked among women born between 1945 and 1960 and then decreased, as observed in Hungary.

North America
Overall, cervical incidence and mortality in the USA have declined for many decades in both black and white populations (Figure 60); this has been attributed to the effect of cytological screening programmes countering any increase due to changes in risk factors (Devesa *et al.*, 1989). Increases at younger ages have not been observed in white or black women (Devesa *et al.*, 1989; Wang *et al.*, 2004).

In British Columbia, Canada, the age-adjusted incidence rate of squamous-cell cervical cancer was 28.4 cases per 100 000 woman-years in 1955, before the large-scale popula-

tion-based centrally organized screening programme, and decreased to 6.4 in 1985 (a 78% decrease) (Boyes *et al.*, 1981; Anderson *et al.*, 1988). The corresponding mortality rate decreased from 11.4 deaths per 100 000 woman-years in 1958 to 3.1 deaths in 1985 (a 72% decrease). The lifetime coverage of cytological testing was estimated at 85% from 1970 onwards.

Although screening for cervical cancer commenced in North America towards the middle of the 20th century (in British Columbia, Canada, in 1949), there was a concomitant decline in mortality from the disease that initially did not seem to be associated with screening (Kinlen & Doll, 1973). Therefore, studies were initiated in the USA and Canada which attempted to evaluate the association between the extent of the decline in mortality from cancer of the cervix and the intensity of screening (Cramer, 1974; Miller *et al.*, 1976). In both countries, a strong association was found when regional declines were analysed in relation to screening data from various sources. In Canada, the cytology data were derived from a national survey and the mortality data were rates among women aged 30–64 years, the ages at which mortality was expected to be most strongly associated with screening (Miller *et al.*, 1976). Mortality from cancer of all parts of the uterus was used, as the extent to which deaths were attributed to cancer of the uterus NOS varied across the country and with time. The association between reduction in mortality and screening was strong at the national level for declines from 1960–62 to 1970–72, and at the census district level, and it was demonstrated that the decline was not explained by census-derived risk factors. In a further analysis, Miller *et al.* (1981) showed that the decline was not explained by changes in hysterectomy rates.

Subsequently, Miller (1986) re-examined the correlation between mor-

tality rate and screening intensity in various parts of Canada for later time periods. Although he found consistent negative correlations between screening intensity and mortality rate at different points in time, he did not find consistent correlations with mortality reduction during time periods after the 1960s. Problems the author noted with this approach were possible changes in the underlying incidence of cervical cancer and the marginal effect expected with marginal increases in screening activity over time once a certain level of activity had been established.

Australia and New Zealand
Although cervical cancer incidence rates in Australia (Figure 61, New South Wales) and New Zealand have not greatly decreased (Coleman *et al.*, 1993), the mortality rates have been clearly declining in Australia for many decades; in women aged under 35 years, decreases have been seen since the mid-1980s in incidence and from around 1990 in mortality. Some increases in mortality rates during the 1970s were noted, notably in younger women (Armstrong & Holman, 1981), and a more recent study observed continuing period-specific declines in incidence and mortality from 1972 to 1996, alongside increasing rates in successive generations (Taylor *et al.*, 2001). Increasing cohort-specific risks in women born in the late 1930s in New Zealand were reported (Cox & Skegg, 1986), but not confirmed later (Cox & Borman, 1994).

Japan
Although incidence and mortality from cervical cancer in Japan have been reported to be falling for many decades (Figure 62) (Coleman *et al.*, 1993), there is evidence of some increase (particularly in mortality) during the 1980s and 1990s in women aged under 35 years. In a study of cervical cancer incidence in Miyagi Prefecture during 1959–87, an age–period–
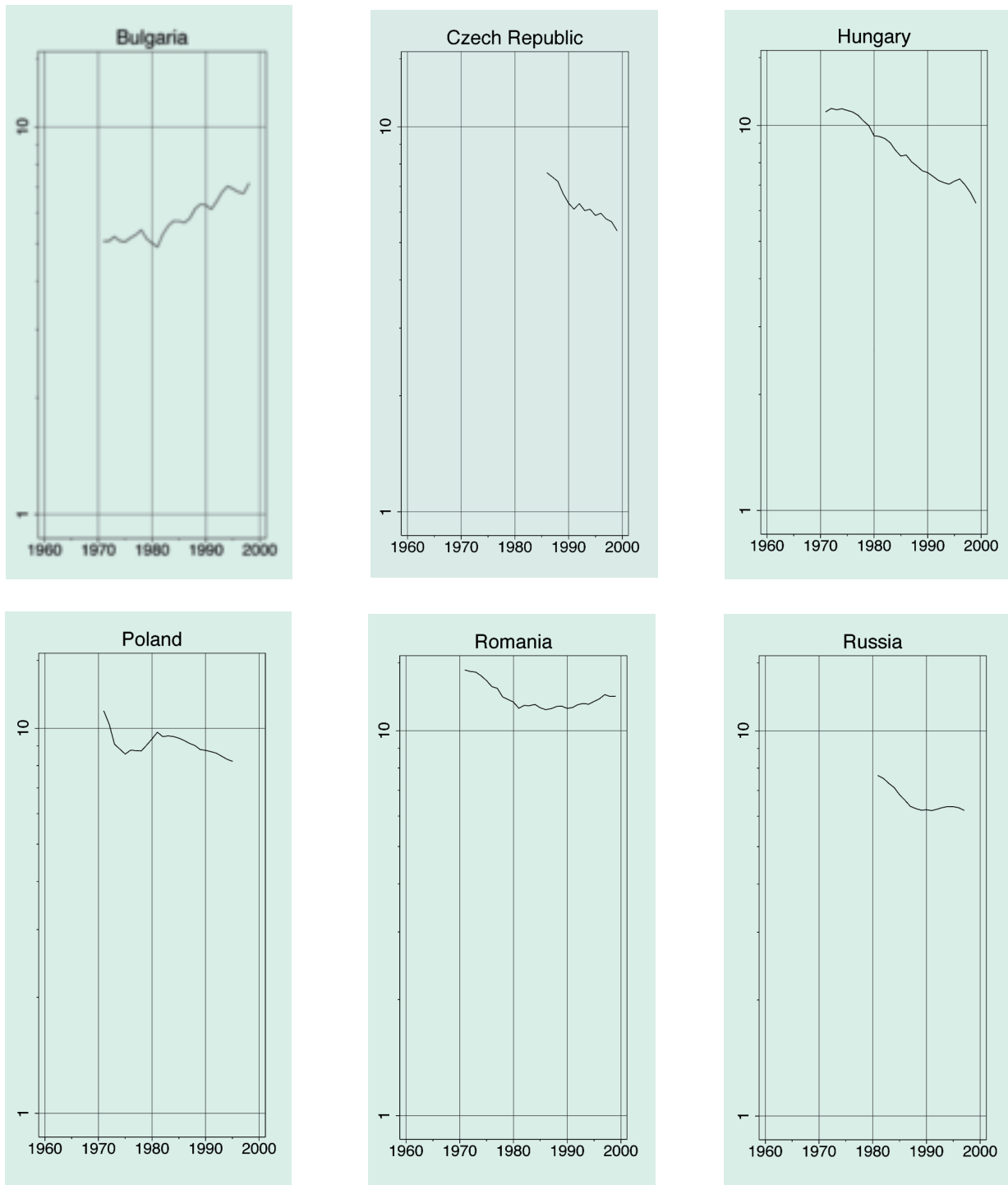
Figure 59 Age-standardized mortality rates of cervical cancer in Bulgaria, the Czech Republic, Hungary, Poland, Romania and the Russian Federation, ages 0–85+
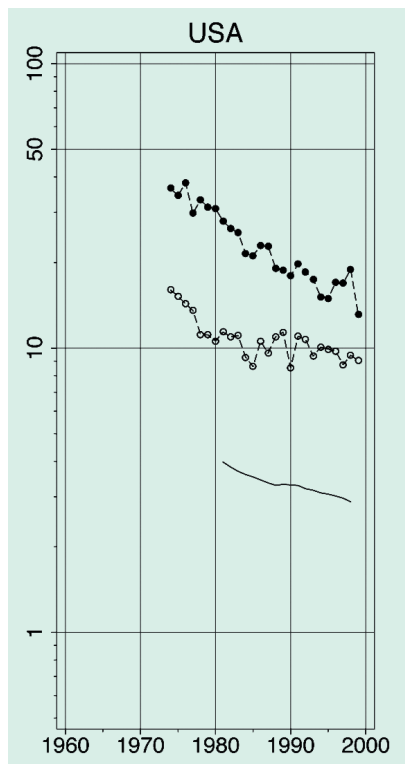
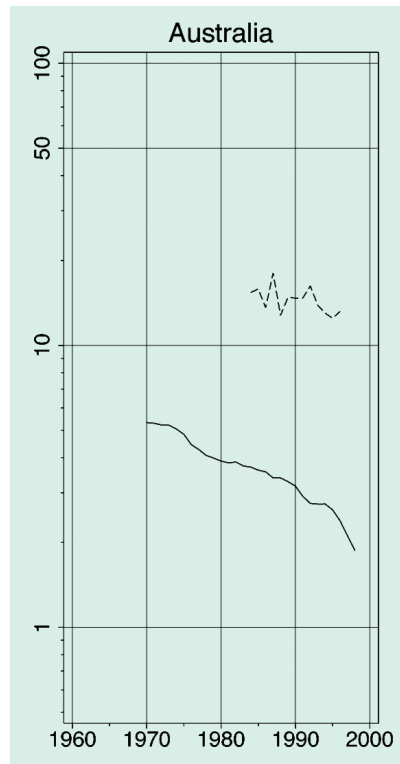**Figure 60** Cervical cancer incidence (-----, ● black, ○ white) and mortality (——) trends in the USA, all ages

Figure 61 Cervical cancer incidence (----) and mortality (——) trends in Australia, New South Wales, all ages
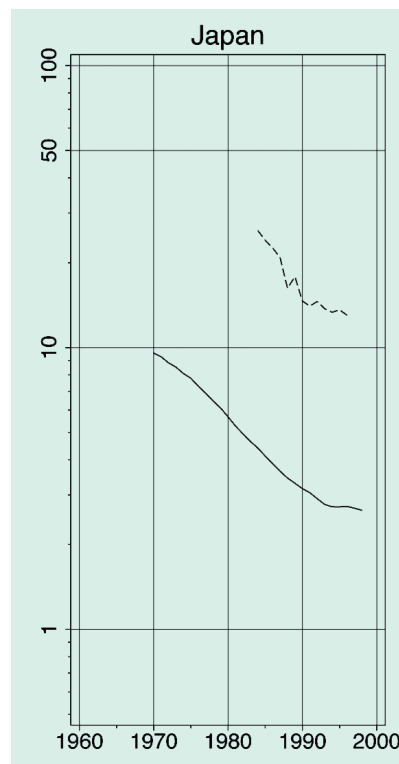
Figure 62 Cervical cancer incidence (----) and mortality (——) trends in Japan, all ages
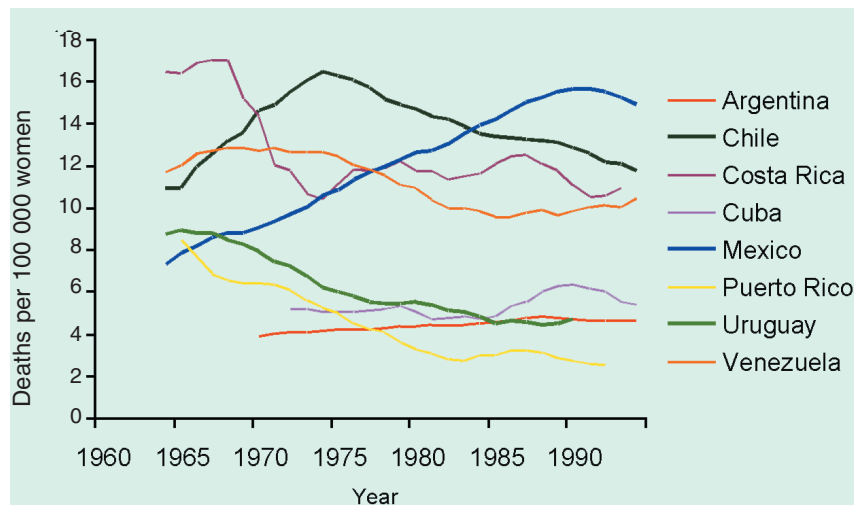
Figure 63 Annual cervical cancer mortality rates (per 100 000) in selected Latin American countries, age-adjusted to the world population, 1960–95. Five-year moving averages

cohort model showed that risk had decreased in recent periods and in younger generations of women (Minami *et al.*, 1996).

Time trends in developing countries
There is limited information on time trends in cervical cancer in developing countries. In general terms, rates of incidence and mortality have been relatively stable or shown rather modest declines (Sankaranarayanan *et al.*, 2001). The absence of the declines in incidence and mortality that have been observed in high-resource populations probably reflects the lack of screening programmes, or, where they exist, the low population coverage and poor quality of cytology (Lazcano-Ponce *et al.*, 1998).

Figure 64 Age-specific incidence rates of cervical cancer in successive time periods
a. Puerto Rico; b. Cali, Colombia

Latin America

In contrast to most developed countries, mortality due to cervical cancer in Latin America increased between 1975 and 1985 (Restrepo *et al.*, 1993). A later analysis (Robles *et al.*, 1996) showed almost no significant downward change in mortality in Latin American countries between 1960 and 1993.
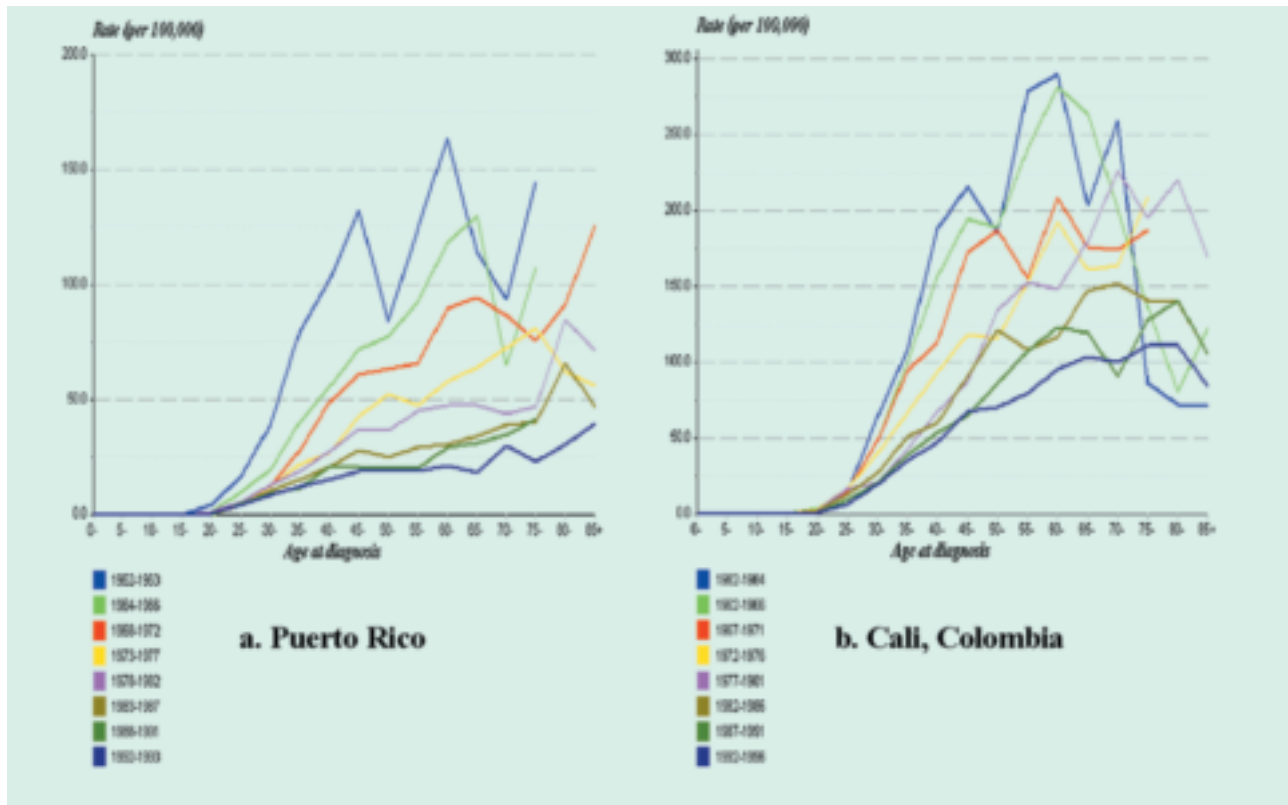
Figure 63 shows trends in age-adjusted cervical cancer mortality in eight Latin American countries between 1960 and 1994. In Puerto Rico, with rates similar to those of Mexico, Venezuela and Uruguay at the beginning of the period, a persistent decline has been observed, that gave it, by the end of the 1990s, the lowest risk in the region. This decline parallels the introduction of a screening programme (Robles *et al.*, 1996), the effect of which can be seen in the progressive decline in age-specific rates, especially in the middle of the age range (30–69), where screening should have the highest effect (Figure 64a). In Cali, Colombia, a decline in the incidence of invasive carcinoma was accompanied by an increase in registrations of carcinoma *in situ* following the introduction of a screening programme in 1967 (Figure 64b) (Aristizabal *et al.*, 1984). In Chile, Costa Rica, Cuba and Mexico, very limited changes in mortality from cervical cancer appear to have followed the introduction of screening. Mortality increased from 1965 onward in Mexico, where a national cervical cancer screening programme was initiated in 1974; although a slight decreasing trend has been observed since the 1990s, the risk remains among the highest in the region. In Costa Rica, cytology testing has been available nationwide to women aged over 15 years since 1970, but mortality and incidence have remained almost unchanged (Herrero *et al.*, 1992). In Cuba likewise, the national screening programme was judged to have had no impact on either incidence or mortality in the period 1980–94 (Fernandez Garrote *et al.*, 1996). In Chile, mortality rates increased steadily between 1960 and 1975, and then began to decrease, although rather slowly. This decline has been modest despite the operation of an organized screening programme since the early 1970s
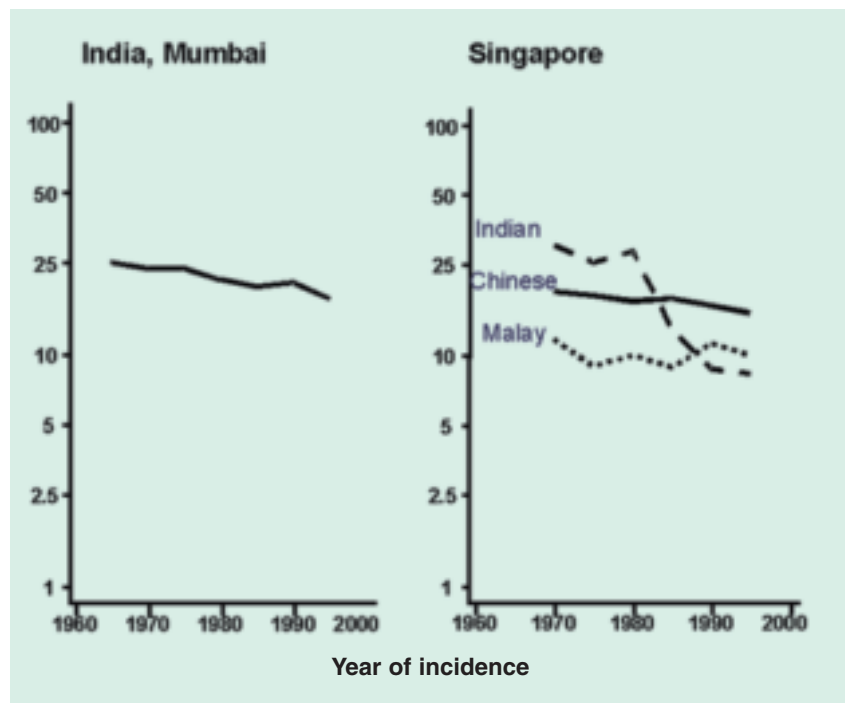
Figure 65 Cervical cancer incidence trends in Mumbai, India and Singapore (Source of data: Cancer Incidence in Five Continents)

(Taucher *et al.*, 1996; Sankaranarayanan *et al.*, 2001). However, the proportion of cancer of the uterus NOS has steadily decreased from almost 50% at the beginning of the 1960s to around 10% in the 1990s, and this would have masked some of the decline in mortality from cancer of the cervix, as described above.

Asia

Figure 65 shows trends in cervix cancer incidence reported by the cancer registries of Mumbai (India) and Singapore. Declines in incidence are relatively modest (except for the Indian population of Singapore, among which the age-standardized rate declined from 29.8 per 100 000 in 1968–72 to 8.2 in 1993–97). In contrast, dramatic declines in cervix cancer in China have been reported. The age-adjusted incidence in Shanghai fell from 26.7 to 2.5

per 100 000 between 1972–74 and 1993–94 (Jin *et al.*, 1999) and mortality rates have fallen dramatically, especially in urban populations, although the trend has reversed recently in younger women (Yang *et al.*, 2003). The declines have been attributed to cytological screening, treatment programmes and improved female genital hygiene, while the increased rates among younger women may reflect changing economic circumstances and sexual habits leading to a greater prevalence of infection with HPV and other agents (Li *et al.*, 2000).

Africa

There are very few data on time trends from Africa. In Bulawayo, Zimbabwe, the frequency of cervical cancer increased significantly during the period 1963–77 (Parkin *et al.*, 1994). Mortality data from South Africa sug-

gested some increase in rates for the 'coloured' population between 1949 and 1979, but little change in the black population from 1964 to 1977 (Bradshaw & Harington, 1985). After about 1980, the mortality in the 'coloured' population remained more or less constant, while in the white population, mortality declined from the mid-1960s (Bailie *et al.*, 1996). The difference was ascribed to the availability of screening services, particularly for older women.

In some registry series, recent incidence rates appear to be higher than earlier ones. In Kampala, Uganda, for example, there has been a significant increase since the 1960s (Wabinga *et al.*, 2000). On the other hand, there seems to have been little change in the recorded rate in Nigeria; it was 20.9 in 1960–69 and 19.9 in 1998–99 (Parkin *et al.*, 2003).

Caveats in the evaluation of time trends in relation to intensity of screening

Trends in cancer incidence and mortality are a complex phenomenon to study, having substantial limitations and potential errors associated with them (Saxen, 1982; Muir *et al.*, 1994). In addition, there are specific issues that concern the interpretation of time trends of cervical cancer, including changes over time in the proportions of deaths certified as uterus NOS and in the prevalence of hysterectomy (see Chapter 1).

The effect of screening is difficult to separate from the effects of other factors influencing rates of cancer diagnosis or death. For example, cervical cancer mortality rates were declining in North America before widespread screening was introduced, and the rate of decline changed little over the period 1946–74 despite considerably increased screening activity (Gardner & Lyon, 1977). This has also been noted in other parts of the world

(e.g., Miller, 1999; Arbyn & Geys, 2002). Some studies have found underlying risk to be greater for recent-born cohorts (e.g., in Belgium, Arbyn & Geys, 2002; England and Wales, Parkin *et al.*, 1985). The increased incidence in young Finnish women in the 1990s may also indicate changing risk, although some inadequacies in screening might also be responsible (Anttila *et al.*, 1999). Because screening has been in use for decades in many parts of the world, especially in developed countries, it is difficult, if not impossible, to accurately estimate the risk of cervix cancer in its absence, in order to estimate the true impact of screening (van Ballegooijen *et al.*, 2000). Further difficulties in interpreting trends result from a lack of information on the extent and quality of screening, particularly when a large proportion of tests are outside organized programmes.

Age, period and cohort effects for squamous-cell carcinoma

Trend studies generally fail to distinguish adenocarcinomas from squamous-cell carcinomas, although their etiology may be rather different, and their susceptibility to detection by cytological screening certainly is (Mitchell *et al.*, 1995b, 2003). In an attempt to further evaluate the effectiveness of screening, the trends in incidence of the squamous-cell carcinoma by age at diagnosis, period of diagnosis, and birth cohort have been examined (Bray *et al.*, 2004) using an age–period–cohort model (Case, 1956; Holford, 1983; Clayton & Schifflers, 1987a,b).

An examination of cancer rates according to birth cohort may provide insight into the nature and intensity of disease-correlated exposures that may vary across successive generations. Cohort effects may relate to birth itself, or may appear to be related to birth only as a result of influences that are shared in the same group as they age together. Temporal changes in environ-

mental risk factors tend to affect particular generations of individuals in the same way as they age together, and are more likely to exert particular influence on earlier stages of carcinogenesis.

Cancer rates by time, on the other hand, may act as surrogate measures of events that quickly change incidence or mortality with the same order of magnitude in all age groups under study. These effects may be the result of planned interventions that act at later stages of carcinogenesis, such as new therapies that improve survival in all age groups. More frequently, they are due to influences that artificially raise or lower the number of observed events (e.g., changes in classification or improvements in diagnostic procedures).

It is likely that any major effect of a general screening policy will be more visible in the period than in the cohort parameters. Such an interpretation is crude and obviously subject to uncertainties; thus, a screening policy may focus on a narrow window of ages and be of short duration, corresponding to a cohort.

Age is a powerful determinant of cancer risk, since it parallels the cumulative exposure to carcinogens over time and the accumulation of the series of mutations necessary for the unregulated cell proliferation that leads to cancer (Peto *et al.*, 1985).

In presenting the age, period and cohort effects for squamous-cell cervical cancer incidence, the effect of age was fixed *a priori* as a biological constant. Two characteristic age curves that related the time before screening distorted the age–incidence pattern. Here the Gustafsson *et al.* (1997a) proposal was applied and the final choice for each population took into account the credibility of the curves from a biological point of view and empirical evidence that the subsequent period and cohort effects were in reasonable agreement with the observed trends.

Figure 66 provides estimates of squamous-cell carcinoma trends from age–period–cohort models for women aged 30–64 years. In Finland, the declines observed since screening was introduced in the 1960s have recently reversed, rates having steadily increased in cohorts of women born after 1945 (Figure 66a) and diagnosed in the 1990s These model-based estimates are consistent with the observed overall time trends (Figure 56) (Anttila *et al.*, 1999). Similar declines in period trend and fluctuation in cohort parameters are seen in Sweden (Figure 66b), although notable changes in rates in younger generations are not clear in the observed trend (Figure 57) (Bergstrom *et al.*, 1999). In England, increasing rates are seen in generations born after 1935 (Figure 66c). In the observed trends, a deceleration in the rise has taken place among very recent generations (Vizcaino *et al.*, 2000). The period parameters for England have reversed since the late 1980s, a finding which is consistent with the overhaul of the screening programme from 1988 (Walker *et al.*, 1998; Quinn *et al.*, 1999). In Estonia, where little screening has taken place (Aareleid *et al.*, 1993), the period parameters have no trend, and there have been clear cohort-driven rises in women born since the mid-1930s (Figure 66d). In the USA, there are clear and uniform cross-sectional declines in period parameter, observed in the rates from the 1970s in both black and white women (Figure 60). Rates were relatively stable among successive generations of white women (Figure 66e) and steadily decreasing cohort trends among black women (Figure 66f).

In conclusion, the age–period–cohort modelling seems to confirm what is known on screening activities, effectively summarizing the data, as well as shedding additional light on the effects of etiological exposures.
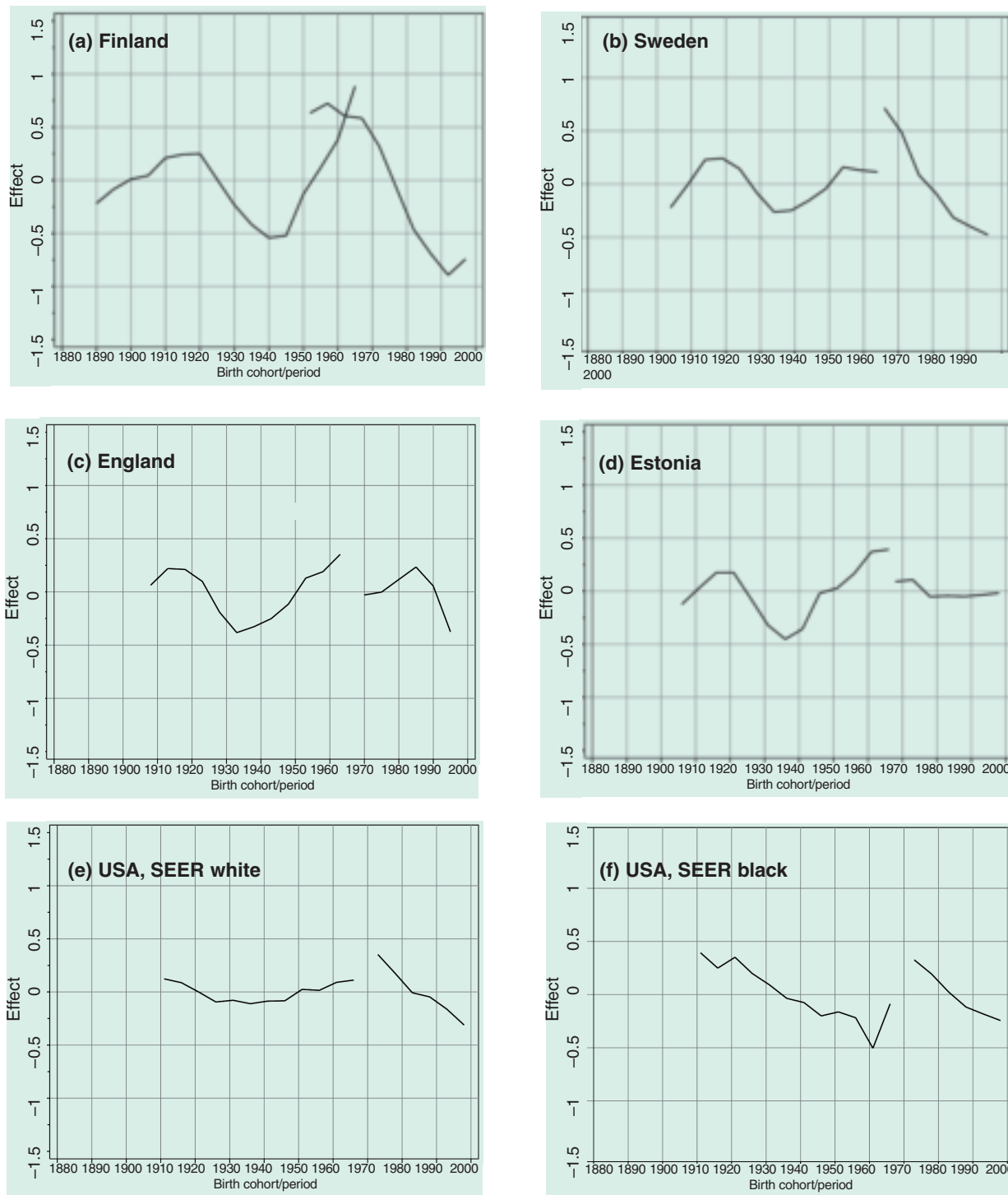
Figure 66 Estimates of squamous-cell carcinoma trends from age–period–cohort models for women aged 30–64 years.
Left-hand curve: cohort parameters. Right-hand curve: period parameters

Use of modelling in the design and evaluation of screening

Statistical models have been developed to explore the effect of screening test, policy and programme characteristics on the expected reductions in incidence and mortality (and derivative quantities such as years of life saved). These have led to improved understanding of the relative importance of various screening parameters, which in turn has made it possible to infer what changes in screening programmes might be most effective. The quality of the models has improved over time as the underlying parameters (natural history, test sensitivity, etc.) have become better understood. The models have also become more widely used, as the contribution of the sophisticated methodology has become better appreciated and the statistical techniques more widely disseminated. As with any model, they depend on the availability of data and on the accuracy of the assumptions.

The pooling of several case–control and cohort studies (IARC Working Group on Cervical Cancer Screening, 1986) was used to estimate the reduction in incidence in a cohort of women as they age from 20 to 64 years under different assumptions of the ages of testing and its frequency (WHO, 1986). The results have been widely quoted and used as input data in various models.

In the absence of direct observations, models can examine the influence of variation in the underlying risk of disease over time or between population subgroups, and in the quality of screening procedures, on the effectiveness of screening as measured by incidence. Simulation models have been developed. These use observed data on the natural history of the disease, screening test performance and effectiveness of different options for treat-

ment of precancerous lesions, and can allow for variation in risk, accessibility, compliance and feasibility. The early models of this type were used to examine the relative effectiveness of different programmes and were relatively simple computer simulations (Knox, 1976; Eddy 1980). More complex models use Monte Carlo simulation methods to provide greater flexibility and more realistic simulation of disease natural history (van Oortmarssen *et al.*, 1981; Goldie, 2002). Gustafsson and Adami (1992) developed a differential equation describing natural history based on the use of computerized identification techniques.

The main findings derived from these models were that, with increasing numbers of tests, the marginal gains become smaller with each additional test (or unit cost). With few tests, the optimal age to start screening is around the age of 35 years, and as more tests are added to the schedule, the optimum age at start diminishes but less than the addition of years for examination at older ages. Attendance, test sensitivity and completeness of follow-up, at moderate levels of screening

intensity, improve effectiveness more than increasing numbers of tests.

Van Ballegooijen *et al.* (2000) used the more general MISCAN simulation model to compare the impact of policies and characteristics of screening programmes across Europe on the modelled reduction in life-years lost due to cervical cancer. They did not take into account possible regional variations in, for example, natural history, underlying risk, prognosis or test sensitivity. They estimated reductions from 21% to almost 100% for different screening policies and coverage rates operative in European countries, under the most conservative assumption about round-to-round participation in screening (Table 72). They also estimated the reductions in incidence and mortality for each country in the light of their screening policy and assuming complete coverage. This work is continuing.

Using modelling techniques without simulation, Goel *et al.* (1998) estimated the impact of various potential improvements to screening in Canada on the incidence of cervical cancer. The results suggested that the number of

Table 72. Percentage reduction in life-years lost according to policy and coverage

	Netherlands, Finland	Belgium, France, Greece, Italy, Spain	Germany
Starting age	30 y	25 y	20 y
Interval	5 y	3 y	1 y
Ending age	60 y	64 y[a]	72 y
Lifetime number of tests	7	14	53
Interval coverage (%)	**% Reduction in life-years lost**		
25	21	24	25
50	42	47	50
75	63	71	75
100	84	94	99.9

[a] For France, the stopping age is 65 years
Adapted from Van Ballegooijen *et al.* (2000)

cervical cancer cases could be reduced by 15% if Canada achieved full coverage with three-yearly screening. If, instead, smear quality were improved so that all smears were satisfactory for evaluation (versus 5% inadequacy assumed by the model), cases would decline by about half this amount.

Hakama and Hristova (1997) used age–period–cohort modelling to project mortality to 2017 across the Nordic countries under three scenarios: with no screening, with present levels of screening (and projections based on recent trends) and with improved screening as for Finland, that was termed optimal screening. They estimated that 91% of cervical cancer deaths were prevented in Finland by screening, i.e., could be prevented with optimal screening. Further, they noted that greater declines than observed could have occurred in the Nordic countries other than Finland had the Finnish (optimal) screening practices been adopted. It was predicted that screening would prevent the loss of 10 000 woman years in the Nordic countries in 2010 and that the costs of the health services were less with screening than without, assuming the organization practised in Finland.

Issues in the implementation of screening

The Papanicolaou test was never rigorously evaluated in a randomized controlled trial such as those to which new screening techniques are subject today. Observational data have been used to demonstrate the efficacy and effectiveness of screening in controlling cervical cancer (see Chapter 4 and above) and it would now be unethical to conduct a randomized study in the presence of existing cytology-based screening. Where screening has failed to work, the blame can be laid on the design or delivery of the screening service (Zapka *et al.*, 2003).

Reaching the women at risk

The many organized screening programmes around the world are described in Chapter 3, but much cervical screening is also undertaken spontaneously. In the USA, where screening is opportunistic or spontaneous, about 80% of women over the age of 25 years report having had a test in the last three years (Breen *et al.*, 2001; Swan *et al.*, 2003). In England, where screening is organized, 81.5% of women aged 25–64 are reported as having had a test in the last five years (Statistical Bulletin, 2003). These are clearly very comparable rates. Gustafsson *et al.* (1995) suggested that a screening test can be equally effective whether performed in an organized or opportunistic setting. [The Working Group noted that Gustafsson *et al.* inappropriately considered the prevalence of carcinoma *in situ* and microinvasive cancer in estimating the effectiveness of screening, rather than the incidence of clinical invasive cancer.] However, while opportunistic screening may avoid the costs of the central call/recall bureaucracy, it can be considerably more expensive (Schaffer *et al.*, 1995).

A further issue is whether screening reaches the population at high risk. In the United Kingdom, before organized call/recall was introduced in 1988, only around a quarter of women had had a recent test and these were largely lower-risk women (Farmery & Gray, 1994). In the context of a generally organized and centrally funded health service, opportunistic screening was failing a large proportion of the population and, in addition, cervical cancer rates were beginning to rise, particularly among younger women (Beral & Booth, 1986). Organization of the screening programme in England and Wales raised the coverage rates and led directly to a 42% drop in cervical cancers between 1988 and 1999, after an initial increase in the number of cases diagnosed (Quinn *et al.*,

2001). Following a case–control study, Nieminen *et al.* (1999) concluded that the substantial decrease in the incidence of and mortality from cervical cancer in Finland was due mainly to the organized mass screening that had taken place rather than to any opportunistic testing and that opportunistic testing was far less efficient.

The part of a population that is hardest to reach generally includes many of the high-risk women (Davey-Smith *et al.*, 1994), for reasons that are surprisingly similar despite the different health systems observed. These include socioeconomic deprivation, cultural and language barriers, often being from a minority ethnic group, being highly mobile in residence and not having a 'usual care provider' (Lawson *et al.*, 2000). Many strategies have been employed to attract such women for screening, with varying degrees of success. It has been found that use of nurses to take smears improves acceptance, particularly among deprived women (Baker & Middleton, 2003) and non-medical smear-takers may also be used (National Cervical Screening Programme (NZ), 1998). Even when the initial test and any follow-up required is provided without charge, difficulties in reaching deprived women can persist (Chiu, 2003). Organized call/recall systems are more likely to reach these women, although the accuracy of the register is a key factor in the degree of success (Baker & Middleton, 2003).

Screening is best organized on a population basis as a public health programme. Evidence from the Netherlands illustrates the difficulties in organizing cervical screening within a general practice setting (Hermens *et al.*, 1998). Even when professional thinking on policies is clear, external assistance is required to bridge the gap between policy and effective delivery of the programme (Hanselaar, 2002).

Most cervical cancer screening now takes place in the developed world, although 80% of the cases are found in developing countries. The scarcity of the skills and resources required to report cervical cytology in developing countries together with the difficulty of finding and treating women have led to interest in investigating alternative techniques for cervical screening in these areas, such as visual inspection of the cervix (see Chapter 4).

Age and frequency of screening

The ages at which screening takes place vary considerably. In some countries, such as the USA, screening is recommended from the age of 21 or three years from the onset of sexual activity. However, in others, such as the Netherlands, screening does not commence until the age of 30 (Coleman et al., 1993). A study in the United Kingdom found that cytology screening was less effective in young women, but grew in effectiveness as women aged (Sasieni et al., 2003). This led to the decision in England to move from a recommended age of 20 years for the initiation of screening to the age of 25 years.

The frequency of screening also varies widely. In the USA, screening has generally been recommended on an annual basis. In several European countries, five-yearly screening is recommended. In England, based on the findings of Sasieni et al. (2003) on the variable effectiveness of screening with age, there has recently been a move to three-yearly screening for women aged 25–49 and five-yearly screening for women aged 50–64 years.

In low-resource settings where organized screening programmes are being developed, optimizing the screening intervals may be less important than ensuring that each woman in the target demographic groups is screened once before any is screened a second time (Suba et al., 2004).

Identifying abnormalities

The process of performing and reporting the original test has a number of distinct phases. The first of these is obtaining the sample. When cervical screening was first implemented in a structured way in British Columbia, Canada, in 1949, the objective was to demonstrate the effectiveness of the technique first reported by Papanicolaou in the 1940s, and the majority of smears were taken by general practitioners during examinations. Most cervical screening still takes place in the primary-care setting, but the nature of the individual who actually performs the test varies from one country to another (Boyes & Worth, 1976)

The sampling device used in the original Canadian system was the Ayre's spatula. This has remained in use to the present day, often in combination with an endocervical brush to ensure sampling of the endocervical canal. Cotton swabs have also sometimes been used [the Working Group noted that this is not an efficient sampling technique]. Extended-tip spatulae, such as the British 'Aylesbury' spatula, have come into use over the last 15 years and more recently plastic brooms. These are almost always used where a liquid-based specimen is to be taken. Buntinx and Brouwers (1996) conducted a meta-analysis looking for any relationship between sampling device and detection of abnormality and concluded that either the extended-tip spatula, a combination of any spatula plus the Cytobrush or cotton swab, or the plastic broom should be used for cervical screening.

The most common screening test remained the conventional Papanicolaou smear until relatively recently, when liquid-based cytology (LBC) was introduced. LBC testing is now used for the majority of cervical screening in the USA (Noller et al., 2003) and this is spreading elsewhere. The United Kingdom is now converting its entire programme (NICE, 2003), as a result of improvements in efficiency due to the dramatic drop in the number of tests reported as inadequate for diagnosis and improved laboratory productivity. LBC may increase the detection rate of cervical screening (see Chapter 4), although studies are generally based on findings at one test, rather than in a population over time, so there is a lack of long-term data (Payne et al., 2000). In the United Kingdom, the introduction of LBC was modelled to be cost-effective, as discussed below (Moss et al., 2003). Changing to LBC facilitates a move to using automated devices to assist in reporting, although few studies have yet provided data obtained with currently available equipment.

The place of testing for high-risk HPV DNA in a cervical cancer control programme is not yet defined. Testing women for high-risk HPV as triage for borderline or ASCUS cytological results is becoming common following the publication of data from studies conducted primarily in the USA (Manos et al., 1999; Solomon et al., 2001; ANAES, 2002; Arbyn et al., 2004). Studies are also being conducted into the possibility of using HPV DNA testing as a primary screening test (see Chapters 2 and 4). Adding HPV testing to cervical cytology allows the interval to be increased for HPV-negative women with normal cytological results (van den Akker-van Marle et al., 2003).

Quality assurance is an integral part of most screening programmes, although in practice it varies from the comprehensive quality assurance programme seen in the United Kingdom to systems which cover the laboratory only, such as the Clinical Laboratory Improvement Amendments (CLIA) in the USA. European guidelines, originally produced in 1993 (Coleman et al., 1993), are currently being revised. Evaluation of screening programmes in the longer term requires monitoring of cervical cancer incidence and mor-

tality rates and comparison of data in the screened population with what might have been seen in unscreened populations (Day, 1986).

Follow-up and treatment of abnormalities

Sasieni *et al.* (1996) calculated that in England, 21% of the cervical cancers with inadequate screening history in 1992 were due to failure to follow up abnormalities according to the then current guidelines. Failure to investigate and treat women with cytological abnormalities and loss to follow-up after treatment are well documented pitfalls in the operation of cervical screening programmes (see Chapter 3). The majority of preinvasive cervical lesions can today be treated under colposcopic guidance and with no or only a local anaesthetic. This has considerably lessened the harm caused by cervical screening compared with the early days when radical surgical techniques were the treatment of choice (Boyes & Worth, 1976). Treatment is now extremely successful, with a complication rate of less than 2% (Luesley & Leeson, 2004). However, due to the risk of recurrent disease, women who have been treated for cervical intraepithelial neoplasia (CIN) are generally recommended to have annual cytological screening for around 5–10 years before returning to a longer cycle.

A systematic review on HPV DNA testing in the follow-up after treatment of CIN indicates that a positive HPV test can pick up treatment failure more quickly than cytology (Paraskevaidis *et al.*, 2004).

Demonstration projects

Each population to which screening will be applied has different characteristics, priorities and health systems. In order to determine the optimal service design and delivery for a given population, a demonstration project should be

undertaken (Miller *et al.*, 2000). This should consider the feasibility of the proposed arrangements and testing of those arrangements in practice.

Hazards of screening programmes

Screening is an unusual medical intervention in that it is an "investigation, which does not arise from a patient's request for advice for a specific complaint" (McKeown, 1968). While there are excellent data supporting the implementation of mass cervical cancer screening programmes, there are also negative consequences of screening large numbers of healthy women in order to prevent significant disease in a few. These include:

- Psychological consequences of a positive screening result, with increased anxiety and fear among women;
- Misunderstanding by women and health-care providers of the meaning of a positive screening test, such that a positive result is interpreted as a 'cancer diagnosis';
- Misunderstanding by women and health-care providers of the meaning of a negative test as implying no risk rather than low risk for cervical cancer, which may lead to underinvestigation of symptoms;
- False positive screening results leading to unnecessary interventions, with both human and financial cost implications;
- False negative screening results giving false reassurance;
- Overtreatment of preinvasive lesions that left alone would neither progress nor cause any clinically significant disease, particularly as there are still no reliable markers to determine which high-grade cervical cancer precursors will progress

to cancer or will remain clinically insignificant;
- Complications of treatment such as cervical stenosis, cervical incompetence and infertility, as well as the results of more radical therapies, such as hysterectomy, with a range of potentially negative sequelae related to the surgical intervention;
- Opportunity costs to the health-care system of introducing a screening programme;
- Impact of incidental findings during screening.

Psychological consequences of screening

There have been few studies directed specifically at evaluation of the psychological impact of participation in a cervical cancer screening programme. In such programmes, women who are well and asymptomatic are required to undergo a gynaecological examination, which for many women is uncomfortable and experienced as a relatively invasive procedure in a private and intimate part of their bodies. After a delay of varying intervals, depending on the quality of the screening service, the woman receives her result. Approximately 1–10% of all smears are considered abnormal. In most screening services, low-grade cervical abnormalities are managed by repeated testing at defined intervals, while for high-grade abnormalities, women are referred for colposcopic evaluation, another uncomfortable and invasive procedure. Both approaches may cause significant anxiety in women (Marteau *et al.*, 1990).

The notification of an abnormal result may cause anxiety and fear among women. For many, the concept of a 'precancerous' lesion is difficult to grasp and the assumption may be that they either have or are at great risk of having an established cancer (Posner & Vessey, 1988). Despite the markedly improved

outcome after treatment for cancer in the past 20 years, many women still equate a 'cancer diagnosis' with a 'death sentence' (Greer, 1985). The word precancer itself causes anxiety, because what women hear is the word 'cancer' without the qualification of the medical meaning that it is a precursor lesion that may never develop into an invasive lesion (Kavanagh & Broom, 1998).

In addition, in many colposcopy services there is considerable delay for an appointment and the waiting period may be associated with acute anxiety, particularly if the woman believes that she has cancer, even though, due to the long latent period in the natural history of cervical cancer, there may be no clinically significant consequence of this delay.

Posner and Vessey (1988) used a semi-standardized interview to study 153 women from the time they were referred to the clinic for colposcopy to after their final check-up. Of these women, 65% described feeling 'worried or alarmed' after receiving notification of their abnormal test and 27% used words such as 'shocked', 'stunned' and 'devastated'. Women's anxiety was related principally to the belief that the positive result implied cancer and death. Further, even after appropriate treatment, 35% of the women still felt afraid of the possibility of cancer and 43% worried about recurrence of disease. Women also reported having a different view of their bodies and a different attitude to sex after a positive test.

There is very little good information on whether women who attend colposcopy clinics after an abnormal test result differ in perception of cancer risk from those who fail to attend, as suggested by Posner and Vessey. Funke and Nicholson (1993) found no difference in such risk perception between women who attended for colposcopy and non-attenders. Lerman *et al.* (1990) found that women who had

been screened in the past three years had less fear of cancer than those who had not been screened in the same period. McKee *et al.* (1999) found no difference between women compliant with colposcopy attendance with regard to fear of cancer compared with those who were non-compliant. They also found no difference in the perception of the gynaecological examination as embarrassing between attenders and non-attenders at colposcopy clinics.

In evaluating women's responses to an abnormal test result, factors other than fears associated with cancer and death also need to be taken into account. For instance, McKie (1993) found that women made negative links between cervical cancer and sexual promiscuity. The epidemiological findings that having multiple sexual partners or a partner with multiple partners increases the risk of cervical cancer may be interpreted by women with abnormal tests as implying that they or their partners have been promiscuous; the comment has been made that 'their character, as well as their cervix, is smeared' (McCormick, 1989).

In most screening programmes, information given about screening is aimed at achieving high uptake of screening, in keeping with the well documented benefit of wide coverage on reduction in cervical cancer. However, giving information that emphasizes only the positive aspects of screening may have negative consequences for some women who feel let down by the screening process, particularly women who receive false negative or false positive results. A belief that screening gives full protection may in itself have negative consequences and it is important that women understand that screening will not prevent all deaths related to cervical cancer. In addition, many screen-positive women will be treated for an asymptomatic condition that, if left

undetected, would never have progressed to a clinically significant lesion.

Concern has been raised that giving realistic information to the public, including explaining that certain individuals can suffer adverse outcomes, can have a negative impact on uptake of screening, and brings up the issue of the 'public good' versus 'individual autonomy'. Wardle and Pope (1992), commented that attention to the psychological costs of screening had lagged far behind the technical and organizational aspects of screening services. Research into this qualitative aspect of screening has suggested a substantial toll of emotional turmoil, but most studies have been uncontrolled and involved subjective evaluation.

Unnecessary treatment, overtreatment and adverse consequences

The appropriateness of screening for the prevention of cervical cancer should be seen as a balance between the beneficial health effects and the adverse effects and costs of intervention. Unnecessary referrals and diagnostic and therapeutic procedures are often cited as the major adverse effect of cervical cancer screening, although few studies have attempted to quantify them. The Dutch Evaluation Commission (Evaluation Commission Cervical Cancer Screening, 1988, reported in Van Ballegooijen *et al.*, 1990) evaluated the amount of diagnostic and treatment procedures induced by cervical cancer screening prospectively and in relation to mortality reduction using data from the Dutch screening programme. A model-based analysis led to the following estimates: (1) a mean duration of preinvasive disease of 17 years, with the shortest at older ages; (2) a regression rate of preinvasive disease of 60% on average, with the highest at young ages;

and (3) a sensitivity of cytology of around 70% for CIN 3. The false positive rate of cytology was assumed to be 0.4%. The group calculated that for five invitations for screening among women aged 37–70 years at eight-yearly intervals, 13 deaths were avoided per million women per screening year. Each death avoided is balanced by 2800 preventive tests, nine women referred for a gynaecological assessment and four for minor treatment procedures (e.g., conization of the cervix). Increasing the invitations to 25 from five would avoid 27 deaths per million women per screening year, but would require 7300 preventive tests, 22 referrals to gynaecology and eight minor treatment procedures. These data clearly showed that more intensive screening greatly increases the need for intervention with diminishing returns for the extra efforts required.

CIN lesions have been treated using a variety of ablative and excisional techniques over the past 40 years (Martin-Hirsch et al., 2004), with each method having its own range of complications and consequences (although the same efficacy). Ablative techniques rely on a histological diagnosis provided by colposcopically directed punch biopsy, which may both undercall (leading to missed diagnosis of microinvasive cancers and under-treatment of these conditions) or overcall (leading to over or unnecessary treatment of the cervix). The most widely used ablative techniques include cryotherapy and laser therapy (see Chapter 1).

In addition to the potential for unnecessary interventions, the widespread use of excisional procedures to treat preinvasive lesions of the cervix may have led to considerable overtreatment. Data on negative histological findings in tissue obtained during loop electrosurgical excision procedure (LEEP) have been reported

in a number of studies using a 'see-and-treat' approach, i.e., treatment of CIN after a colposcopic diagnosis without prior histological confirmation. Murdoch et al. (1991) reported an overall 41% rate of negative histology after LEEP in a highly selected group of women attending a colposcopy clinic because of abnormal cytology. In women who had had prior histological sampling, negative LEEP histology was found in 43%, compared with 38% of women treated with LEEP on a see-and-treat basis. Of the women who had negative LEEP histology and who were treated on a see-and-treat basis, the majority (53%) had index cytology of CIN 1. As a consequence, the authors cautioned against the see-and-treat approach in women with low-grade referral cytology.

In a retrospective analysis of LEEP performed at a colposcopy clinic in South Africa (Denny et al., 1995), 18% of LEEPs performed after prior histological sampling (21 out of 116) yielded histologically negative findings compared with 14% of those treated on a see-and-treat basis (16 out of 114). The women were referred to this colposcopy clinic as a result of persistent LSIL (2–3 LSIL cytological results over 12–18 months) or one result of HSIL or suspicious of malignancy. An additional finding in this study was that 25% of punch biopsies were falsely negative; the authors emphasized that a punch biopsy is only as reliable as the colposcopist's ability to identify the most abnormal area for biopsy.

Rates of negative LEEP histology ranging from 5 to 41% have been reported. All of these studies were performed in colposcopy clinics where women had been referred with abnormal cytology (Prendiville et al., 1989; Luesley et al., 1990; Whiteley & Olah, 1990; Bigrigg et al., 1991; Hallam et al., 1993; Denny et al., 1995). False positive cytology or col-

poscopy, false negative histology in the LEEP specimen and possible complete excision or spontaneous resolution of the lesion after prior biopsy are possible explanations for negative LEEP histology. In addition, some series have reported negative LEEP histology where there has been extensive thermocoagulation preventing a histological diagnosis.

While excisional procedures of the cervix are performed under local anaesthetic in an outpatient setting and complications are considered relatively benign, this is not always the case. In a randomized trial of LEEP, cryotherapy and laser vaporization of the cervix, Mitchell et al. (1998) reported complications from LEEP in 7.6% (10/130) of women treated. The majority of the complications related to bleeding (70%); there was one case of infection, one woman complained of severe pain and required treatment and one woman subsequently developed cervical stenosis.

Ferenczy et al. (1996) reported on 1070 women who underwent LEEP and who returned for post-LEEP follow-up. Complications were recorded in 71 women (7%); 37 had significant intra- or post-operative bleeding, and one required admission to hospital to control bleeding. Seven women developed a purulent vaginal discharge and pelvic pain within one week of treatment. A further 13 women (1.2%) developed cervical stenosis, of whom 11 were over the age of 45 and not on hormone replacement therapy.

Cervical stenosis, while rather rarely reported, is a significant complication of treatment of the cervix and may result in infertility or menstrual complications such as haematometria, making follow-up with cytology and/or colposcopy difficult if not impossible. These complications may necessitate hysterectomy, with all the potential sequelae associated with this more radical surgical intervention. Hysterec-

tomy may also be the result of clinical uncertainty as to the meaning of persistently low-grade cytology or persistently inadequate tests, neither of which can be resolved satisfactorily.

In some women, if the excisional procedure removes large amounts of cervical stroma, a complication of treatment may be cervical incompetence. This has been associated with pre-term delivery due to premature rupture of membranes (Sadler *et al.*, 2004)

In addition to the complications associated with treatment, the risk of persistence or recurrence of lesions after treatment makes long-term follow-up of treated women essential. Reports from both non-randomized and randomized trials of treatment for cervical cancer precursors using a variety of treatment modalities indicate that most treatments have about a 90% success rate (Martin-Hirsch *et al.*, 2004).

Opportunity costs to the health system

Setting up a screening programme designed to detect disease in healthy individuals necessitates diversion of human and financial resources from other health interventions, in particular, treatment of already existing or apparent disease. Thus a screening programme and its hazards and benefits need to be evaluated in the context of the competing health needs of the specific country to ensure that resources are used to the maximum benefit of the entire population.

One feature of cervical cancer screening programmes in low-resource countries with endemic HIV infection is the diagnosis of CIN lesions in as many as 20–30% of HIV-infected women. This may lead to the consumption of scarce health resources for treatment of CIN. However, in countries such as Zimbabwe where HIV infection has become pandemic, HIV-infected women succumb to opportunistic infec-

tions long before invasive cervical cancer arises (Chokunonga *et al.*, 1999). In situations where it is not possible to provide any form of treatment for HIV-related conditions (e.g., treatment of opportunistic infections and/or anti-retroviral therapy), cervical cancer screening may not be a priority.

Incidental findings

While screening is an activity designed for healthy, asymptomatic women, there can be unexpected incidental findings at the time of screening that may cause harm as well as benefit to women and the health system. For instance, the discovery of underlying diabetes due to the diagnosis of diabetic vulvitis at the time of performing the screening test may be of benefit to the woman, but the discovery of a lethal co-existent cancer may only increase suffering without offering any improvement in quality of life, if there is no effective treatment for that cancer.

In low-resource countries where women generally have little access to health care, screening may identify a significant number of women with co-morbid health conditions which it is impossible to manage with the available resources.

One possible incidental finding of particular significance is identification of a woman as HIV-positive. The prevalence of CIN among women infected with HIV is nearly five times that in HIV-negative controls (Chirenje *et al.*, 2002). In low-resource countries where cervical cancer screening is mainly opportunistic, there is a large burden of women harbouring CIN lesions with concomitant HIV infection that has not been identified.

Cervical cancer screening in HIV-positive women

It is estimated that there are now up to 42 million people worldwide living with HIV infection or AIDS. About 70% of these individuals live in sub-Saharan

Africa (UNAIDS, 2003) and more than half of the infected people are women; cervical cancer screening has been widely unavailable in this area up to now.

In many HIV-endemic countries, CIN lesions may be detected at the same time that HIV infection is first diagnosed. The discovery of CIN lesions in an HIV-infected woman may create major psychological and morale challenges, not only to the woman, but also to health workers.

The diagnosis of HIV infection still carries a high level of stigmatization and fear of disclosure of an incurable sexually transmissible disease, with some women being exposed to violent response from their male partners and permanent psychological isolation from the community. In the absence of antiretroviral treatment, some women become suicidal and in this group an added diagnosis of a CIN lesion through cervical cancer screening causes special concern. Linkage of a cervical cancer screening programme to HIV testing may present a new barrier to participation in cervical screening.

Another concern is how to interpret a positive cervical screening test result in the presence of HIV-positivity, bearing in mind that natural progression of both conditions is dependent on availability of effective treatment. Treatment of CIN lesions in HIV-positive women by ablative or excision procedures results in epithelial disruption which can theoretically enhance viral acquisition or transmission.

Health workers treating HIV-positive women for cervical disease should take universal precautions as for all health interventions.

Performance evaluation

The minimal essential elements of a cervical screening programme are: a defined population to screen; invitation to this population to participate;

assessment of coverage; a quality control system; and treatment for test-positive women. The means of achieving these, e.g., population registers or geographical location to define the population; personal invitation letters (call/recall) or mass education to invite women to participate; population-linked cervical cancer registry or sample surveys to assess coverage) will depend on the local circumstances.

The recognition that the effectiveness of screening "is determined by the proportion of progressive lesions that are successfully detected and treated" (Pontén et al., 1995), which depends in turn on screening policy and its implementation, has led to the development of indicators of screening programme performance for routine monitoring. The determinants of this proportion represent the essential elements of a good screening programme (Hakama et al., 1985). They include coverage of the target population (participation); attendance for rescreening; adequacy of smear-taking; quality of interpretation of smears; and follow-up of abnormal results. Specific indicators for these programme performance areas have been developed in the context of cytology and more specifically Pap smears.

The Council of Europe recently recommended all Member States to offer organized screening for three cancers including cervical cancer, and stressed that this should be managed in such a way that the performance can be evaluated fully (Council of the European Union, 2003). Organized screening requires adequate data collection systems to be set up concerning invitation and participation of the target population, registration of screen test results and follow-up of screen positives (Arbyn et al., 1999; Advisory Committee on Cancer Prevention, 2000). Screening databases, including personal records, should be linkable with cancer and mortality registers, in order to allow full evaluation of the programme. This needs to be done with full respect for national legislation. Opportunistic screening systems are in general less cost-effective and do not allow evaluation (Advisory Committee on Cancer Prevention, 2000).

Screening policy

Unlike breast cancer screening, where optimal screening policy (in terms of age range, frequency and modality) has been determined by randomized trials and most programmes follow similar policies, results of observational studies of protection offered by cytological screening have produced estimates of effectiveness for a variety of alternative policies. Thus, the maximal effectiveness of a programme will depend on what policy is adopted. The impact of policy on effectiveness has been modelled recently for European countries by van Ballegooijen et al. (2000) (see Table 72).

Most screening programmes recommend the same screening interval for the entire target age range. However, a recent audit of screening in the United Kingdom suggested that the policy of one screening interval across the entire target age range may not yield optimal effectiveness and recommended screening women aged 25–49 more frequently than older women (Sasieni et al., 2003). This further illustrates how effectiveness can be determined in part by policy.

Many factors influence what policy is adopted. For example, a national workshop in Canada recommended that screening intervals not be raised to three years without the security provided by an information system (Miller et al., 1991). Since most Canadian jurisdictions have not in the past had information systems, this almost certainly resulted in over-screening. Medical legal issues may also influence screening policy (and therefore effectiveness) in some countries. For example, a shorter screening interval is safer than a longer one and a broader age range safer than a narrower one. However, the marginal utility of each extra test diminishes rapidly. This is of particular concern in developing countries.

Screening delivery

It is generally accepted that, for optimal effectiveness, cervical cancer screening should be offered within an organized programme (see Chapter 3). The programme should further include the specific elements as described by Hakama et al. (1985) (see above). Hakama and others have concluded that organized screening programmes, as practised to varying degrees in the Nordic countries and particularly in Finland, are more effective than opportunistic screening activities. This conclusion was based largely on comparison of incidence and mortality trends (e.g., Hakama, 1982; Läärä et al., 1987) and cohort studies, as discussed previously in this volume. A case–control study by Nieminen et al. (1999) suggested that tests in an organized programme may be more effective than those delivered outside the programme, within the same jurisdiction. However, since countries differ greatly in their screening practices as well as in the level of organization, it is difficult to attribute higher effectiveness definitively to a more organized programme.

The premise that a totally organized programme is significantly more effective than a programme based on largely opportunistic screening with some central coordination and elements of organized screening, such as is practised in many parts of the developed world, has been challenged (Madlensky et al., 2003). Finland may be one of the very few countries in the world whose screening programme

meets all the criteria outlined by Hakama *et al.* (1985). Although it may not be feasible to adopt all these elements of organized screening, many jurisdictions have incorporated some of them. For example, recruitment strategies may be targeted to specific population groups in the absence of a population register that allows personal invitations to screening and this can lead to achievement of low rates of cervical cancer. For example, incidence rates for cervical cancer (uncorrected for hysterectomy prevalence) and trends over the past 30 years are almost identical in different provinces of Canada: British Columbia, which has had centralized cytological screening and a cytology information system for several decades (see, for example, Morrison *et al.*, 1996); Ontario, which has a recent information system including about 80% of the province's screening tests and a programme that sets policy and standards but where smears are taken and read in a totally decentralized system; and Quebec which has no organized programme and no information system but opportunistic screening (see Figure 67). However, unless historical patterns of screening and cancer incidence rates are taken into account, inferences from such data regarding effectiveness are uncertain (see Chapter 4).

On the other hand, poor organization of a screening programme can produce poor outcome. In the United Kingdom during the 1970s and 1980s, population coverage was low; low-risk groups were over-screened and the technical quality of screening process parameters was moderate. Imple-mentation of call/recall systems and targeted rewards for primary care providers, achieving high levels of coverage among eligible women (Patnick, 2000), resulted in a substantial decline in incidence and mortality in all age groups of the target population (Sasieni *et al.*, 1995; Quinn *et al.*, 1999; Sasieni & Adams, 1999, 2000).

There are, however, several common problems with opportunistic screening versus organized screening, and opportunistic screening should therefore be discouraged:

1. It is less cost-effective (see below)
2. Hard-to-reach women are less likely to be adequately screened
3. In many settings, especially in developing countries, there is disproportionate representation of women who are in contact with the health-care system for other health interventions such as reproductive care, so that those in some age groups are inadequately screened (Were & Buziba, 2001)
4. It can create sporadic work flow, which can lead to reduced proficiency, etc.
5. It can result in greater chances of harm due to over-screening.
6. It may be difficult to ensure quality.
7. Screen-positive women may not have easy access to diagnostic and treatment services.

Performance indicators

An integrated information system or a set of systems that can be linked as required is recommended as the ideal support for performance monitoring; such a system can also support programme operation (Miller, 1992). These systems should permit identification of each woman as well as each test and link them. A model for a comprehensive information system is shown in Figure 68.

For performance monitoring, the system should ideally contain a screening database including results of cytology and follow-up (colposcopy, histopathology, treatment) with periodic linkage to a population register, tumour registry, mortality file and hysterectomy data. However, even in areas where population registers and/or other files do not exist or are not accessible, information systems can be developed that permit estimation of many indicators. Others can be estimated periodically by special studies. A number of indicators are based on negative test results. This generally assumes exclusion of negative results for women who are under special surveillance (e.g., following colposcopy, previous positive history, etc.).

Table 73 outlines performance indicators coinciding with the determinants identified by Pontén *et al.* (1995) and Hakama *et al.* (1985), along with the programme area they are designed to specifically evaluate and required data where relevant. These have been adapted from Coleman *et al.* (1993), who proposed a menu of specific indicators with targets for the Europe Against Cancer Programme.

All indicators should be evaluated, reviewed and published annually. Some of the indicators require multiple sources of information, sometimes linked at the individual level. Where this is not possible, alternative methods, such as periodic special studies (e.g., the health and fertility surveys noted in Chapter 3), should be used.

Participation (or coverage)

The participation rate is the proportion of eligible women in the target population who participate in screening within the time interval specified by local screening policy. It can alternatively be defined in terms of some chosen period of time (e.g., tested in the last three years). Its estimation requires, at a minimum, counts of screened women in the target age range and of the target population. If estimates of the prevalence of hysterectomy are available, they should be used to reduce population counts to reflect more accurately the population at risk. Women with only inadequate smears in the interval should not be counted as having been screened in that
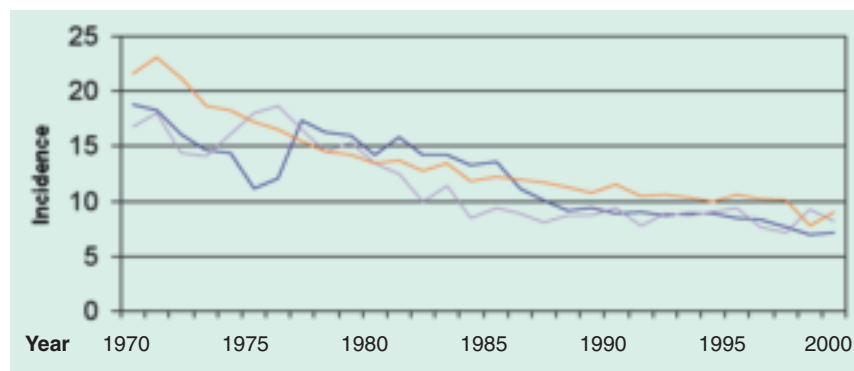
Figure 67 Cervical cancer incidence in three large provinces of Canada, with different intensity of programmatic components, 1970–99
(——) Quebec, (——), British Columbia, (——) Ontario

interval. Participation should be evaluated according to age group, geography and other locally relevant indicators (e.g., health-care provider, ethnicity).

The effect of participation in reducing mortality and incidence has been demonstrated descriptively. For example, in the United Kingdom, Quinn *et al.* (1999) showed that incidence declined dramatically starting in the late 1980s after the introduction of a call/recall programme which resulted in greatly increased coverage. Miller *et al.* (2000) stated that "… the programme must focus on achieving the highest possible coverage rate. To support this, indicators such as number of women

Table 73. Short-term performance indicators for assessing programme effectiveness and efficiency

Programme component	Measure	Definition	Comments
Attendance for screening	Participation (or compliance) rate[a] in relation to programme policy	Percentage of women in the target population with at least one test within the recommended interval	Adjust denominator for prevalence of hysterectomy. Estimate for age, region and other risk indicators
Adequacy of referral and treatment systems	Compliance with recommendations for follow-up of unsatisfactory smears and positive test results	Percentage of women (or tests) with specific positive (or unsatisfactory) results with follow-up action according to recommendations	'Recommendations' are those set by the programme and include follow-up action (e.g., repeat test, colposcopy) and time to follow-up. Report according to reason for follow up (e.g., inadequate, ASCUS, etc.) and person/institution responsible for ensuring follow-up. Reasons for non-compliance should be noted.
	Compliance with treatment recommendations	Percentage of women requiring treatment who receive it according to recommendations	'Recommendations' include type of treatment and time to treatment. Report and investigate as for follow-up compliance.
Overall	Stage of invasive cancers	Percentage distribution of stage at diagnosis for all invasive cervical cancers in region	Requires population-based cancer registry.
	Interval cancers	Number of invasive cancers diagnosed following negative test result and before next expected screen per 100 000 person years at risk[b]	Person years calculated from date of test to date of diagnosis, date of next expected screen or date removed from population at risk. Reasons for individual cases should be investigated as a kind of audit, along with other cancers that are not interval cancers
Efficiency	Over-screening	Percentage of women with negative test who are screened again before end of screening interval[b]	

[a] Also generally referred to as 'coverage'. However, this is an ambiguous term as it can also refer to the percentage of population targeted for screening.
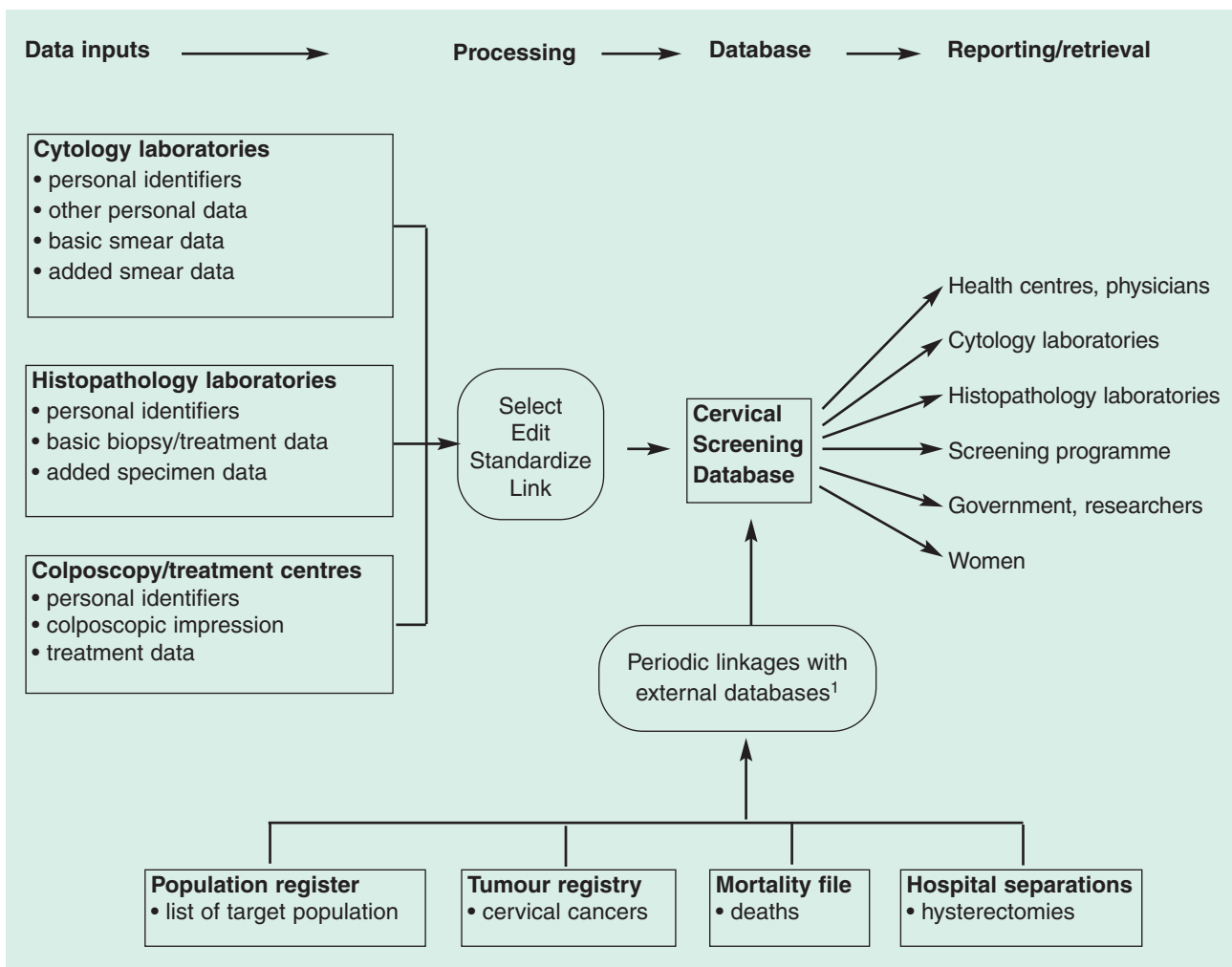[b] May also be expressed per 100 000 negatively screened women

Data inputs ⟶   **Processing** ⟶ **Database** ⟶ **Reporting/retrieval**

Cytology laboratories
• personal identifiers
• other personal data
• basic smear data
• added smear data

Histopathology laboratories
• personal identifiers
• basic biopsy/treatment data
• added specimen data

Colposcopy/treatment centres
• personal identifiers
• colposcopic impression
• treatment data

Select Edit Standardize Link

Cervical Screening Database

Health centres, physicians
Cytology laboratories
Histopathology laboratories
Screening programme
Government, researchers
Women

Periodic linkages with external databases[1]

Population register
• list of target population

Tumour registry
• cervical cancers

Mortality file
• deaths

Hospital separations
• hysterectomies

Figure 68 Model for a comprehensive cervical screening information and reporting system
[1] Periodic linkages would uptake addresses, identify unscreened women, ascertain 'failures' (cancers) and remove women who no longer need screening
From Marrett *et al.* (2002)

screened, as opposed to number of Pap smears done, should be promoted".

Modelling has also shown that participation rate is the most important programmatic determinant of effectiveness. Van Ballegooijen *et al.* (2000) modelled the expected reduction in life-years lost for a number of European countries as a function of screening policy and coverage. They found that differences in coverage resulted in more or less proportional differences in effectiveness, essentially independent of screening policy (see Table 72).

High rates of participation are, however, not sufficient to ensure high effectiveness, if other essential elements of a screening programme are suboptimal. This is evident from the situation in some Latin American countries (see Chapter 3 and this chapter).

In populations where the prevalence of prior hysterectomy is substantial, participation rates will be underestimated, especially in women aged over 50 years, if adjustment is not made. Because rates of hysterectomy vary across time and place, lack of

such adjustment may invalidate comparisons of participation rates over time and between populations.

It is most appropriate to estimate participation based on all screening tests, whether from an organized programme or opportunistic.

Quality indicators

For programmes to be effective, all parts should be quality assured, with indicators for each part of the programme. Furthermore, the programme needs to have access to such quality assurance information and to ensure that it is fed back, along with standards or comparisons, to those providing the service (e.g., primary care providers, laboratories, etc.). Quality indicators should cover test-taking (e.g., inadequacy rates), interpretation (e.g., positive predictive value), treatment and follow-up and programme organization. Acceptable ranges for any such indicators should be specified and values that are outside the range should be investigated. Many of these are discussed in more detail in Chapter 3. Programmes should regularly audit cases of cervical cancer to identify possible shortcomings of the programme.

Modelling of the impact of quality improvements suggests that they have less impact than improvements in coverage. However, in a programme with high coverage, improvements in quality can increase effectiveness. Goel *et al.* (1998) estimated that the impact of reducing the false negative rate from 0.25 to 0.10 (i.e., increasing sensitivity from 75% to 90%), while leaving screening otherwise unchanged, might result in 25% fewer incident cases. As noted by Fahey *et al.* (1995), however, even the lower of these sensitivity values may be higher than is generally achieved. Modelling of effectiveness of screening programmes must use realistic estimates of test quality if it is to provide valid estimates of effectiveness.

Follow-up

Preinvasive (or early invasive) lesions identified via screening must be treated if development of invasive (lethal) disease is to be avoided. Thus, referral for and presentation at follow-up of a positive test result as well as post-treatment follow-up of confirmed precancerous lesions, in accordance with treatment policy, are important. It is also essential that test results be provided to the health-care provider and woman, including specification of the need for follow-up, in a timely fashion.

Follow-up consistent with recommendations

A screening programme should have a policy for follow-up of unsatisfactory tests and positive test results. Policy should specify the action required and the time frame within which this action should take place for specific test results and patient history. Actual follow-up should then be monitored in relation to programme policy. An unsatisfactory test means that a woman has not been adequately screened and the appropriate follow-up recommendation in this situation is for a repeat test. Compliance with this recommendation should also be monitored.

The proportions of women who are followed up according to programme policy (in terms of both action and time frame) should be calculated according to reason for follow-up. Reasons for follow-up failure should be documented.

Computation of indicators of compliance with follow-up for positive test results requires linkage between these tests and follow-up data. This often includes information on colposcopy visits.

Treatment consistent with recommendations

Screening programmes should also have policies regarding treatment of confirmed abnormalities, again specifying both action and time frame.

As with abnormal or inadequate screening test results, the proportions of women receiving adequate treatment should be calculated according to reason for treatment. Reasons for non-compliance should be documented.

Goel *et al.* (1998) estimated that improving efficacy of follow-up and treatment following a positive test result from 0.8 to 0.9 might reduce the number of cancer cases by 2%. The overall impact is relatively small because this improvement affects only women with positive results, which represent a very small fraction of all tests. Pontén *et al.* (1995) estimated that in a programme with fewer lifetime screenings than in the Canadian context, about half of the ultimate reduction in mortality might result from detection and treatment of early invasive disease and half from removal of screen-detected precursors. In such situations, an improvement in the efficacy of follow-up might have greater impact on mortality.

Overall short-term indicators

Because cervical screening can detect asymptomatic invasive disease, an effective programme would be expected to lead to a shift towards more microinvasive and early invasive disease. Thus the major changes in incidence that occurred in Iceland following introduction of screening were among advanced cancers (Johannesson *et al.*, 1982). Evidence for the effectiveness of screening due to early detection of disease can also be obtained by examining trends in survival (see, for example, Adami *et al.*, 1994). However, observed increases in survival can be due to improvements in treatment. In general, a trend in survival is not a good indicator of an effect of screening (see Chapter 1).

Stage distribution of incident cervical cancers

The distribution of all newly diagnosed cervical cancers in the programme area according to stage at diagnosis should be calculated annually. This requires a population-based cancer registry that includes standardized staging information on all or at least a high proportion of newly diagnosed cases. If such data are not routinely available, periodic special studies can be conducted. This indicator is particularly important in areas where screening is not yet well established and screening intervals are relatively long. In such situations, a relatively large part of the effect of screening on mortality will be due to earlier stage at diagnosis.

Interval cancers

Interval cancers are those that arise following a negative test and before the next scheduled screen. These can arise for two reasons: either the previous result was a false negative or the (pre)cancer was not detectable at the time of the previous screen (for example, because it was fast-growing or did not go through detectable preclinical stages). A test repeated every three years on women aged 35–64 years has been estimated to 'prevent' 84–91% of invasive cancers (Day, 1989; Sasieni *et al.*, 2003) if all abnormalities are effectively treated. In the United Kingdom, testing at three-year intervals was estimated to be capable of preventing about 70% of all cancers occurring in women aged 40–69 years, after allowing for cancers that develop in women with positive test results (Sasieni *et al.*, 2003).

The interval cancer rate is calculated as the number of interval cancers per 100 000 person years at risk. Person-years at risk are estimated by summing time from the date of the last negative test to the end of the recommended screening interval or to diagnosis of cancer (or until a woman becomes ineligible due to emigration, death, etc.) for all women having a negative test. Use of this rate to assess programme effectiveness is difficult in regions where screening is well established, because the expected rate of cervical cancer in the absence of screening is unknown. However, the interval cancer rate can be followed over time and compared across programmes with similar screening policies. It may be closely paralleled by the ratio of the number of cancers diagnosed in screen-negative women during the screening interval divided by the number of screen-negative women, expressed per 100 000 women.

Calculation of the interval cancer rate requires knowledge of cancers occurring in negatively screened women. In general, this requires linkage between a population-based cancer registry and the screening test data. In many regions, no cancer registry exists or linkage between screening data and the cancer registry, at least on a routine basis, is not permitted. In such cases, audit studies of interval cancers should be performed.

Screening histories (preceding test and other follow-up results) of all invasive cancers, whether they are true interval cancers or not, should be examined routinely to identify areas where programme improvements may be required.

Indicators of efficiency

There are a variety of parameters that indicate the efficiency of a screening programme. The most important of these relates to over-screening. This can be assessed by the proportion of women with a negative result having a subsequent test before the end of the screening interval. Retention of screened women for rescreening can be assessed by the proportion of screen-negative women returning for rescreen at about the right time. Both of these can be examined in relation to length of interval between tests, region, smear taker, etc. The average number of tests per woman during the recommended screening interval is another indicator of the extent of over-screening.

Process quality indicators

A number of process indicators should be monitored to ensure that screening is operating as it should. Those that are chosen will depend on the issues that are important in a particular setting. These might include elapsed times (e.g., between test-taking and reporting; between reporting of a positive result and follow-up colposcopy), results of laboratory proficiency testing, etc. Acceptable ranges for any such indicators should be specified and values that are outside the range should be investigated.

Economic evaluation and cost-effectiveness of cervical cancer screening

The basic principle of a decision analytic approach is that all consequences of decisions (e.g., individual clinical outcomes, population-based outcomes and costs) should be identified, measured and valued. When a decision analysis formally compares the relationship between the health and economic consequences associated with different public health care interventions, it is considered a cost-effectiveness analysis. The application of economics to public policy does not necessarily mean that less money should be spent, but rather that the use of resources might be more efficient.

Different types of economic evaluation are commonly confused. For example, there are distinct differences between cost-minimization analysis (how much money can be saved?) and

cost-effectiveness analysis (how much health improvement can be gained, per unit expenditure?). The results of a cost-effectiveness analysis are summarized using an incremental cost-effectiveness ratio. In this ratio, all health outcomes associated with a particular strategy (compared with an alternative) are included in the denominator, and all costs or changes in resource use with a particular strategy (compared with an alternative) are included in the numerator. This type of analysis defines the 'opportunity cost' of choosing one clinical or public health approach over another.

Advances in the various technologies and preventive modalities for cervical cancer screening mean that policy-makers in national and international agencies are confronted with various strategies from which to choose. However, there are important differences between developed and developing countries in the policy questions that are most relevant to cervical cancer control. Scarce resources, limited infrastructure and competing health priorities have prevented most low-resource countries from implementing successful cervical cancer screening programmes.

In countries classified as low-income economies (gross national income per capita equal to or less than US $755 in 2000), the key problem is how to implement a sustainable screening programme in the setting of competing health priorities and limited resources. If a cytology-based screening programme is to be introduced, cost-effectiveness modelling using locally derived information about costs and the age–incidence curve for cancer in the population in question, together with internationally accepted data on efficacy, will assist in deciding on the number of screening rounds and the age group to be targeted.

For a developed country where a conventional cytology screening pro-

gramme already exists, information on effectiveness may be available, but whether it is cost-effective may not have been fully established. However, it is possible to assess the likely cost-effectiveness of a new technology in detecting precursor lesions of cervical cancer relative to conventional cytology. For example, England recently used cost-effectiveness modelling as a major consideration in deciding to move to using liquid-based cytology in its programme. This modelling began with the assumption that the effectiveness of new and conventional cytology was the same (Payne *et al.*, 2000). The evaluation of the new technology was then based on costs derived locally from pilot implementation of liquid-based cytology (Moss *et al.*, 2003). The Payne *et al.* (2000) conclusion of equivalence of effectiveness between conventional and liquid-based cytology was based on a surrogate measure of effectiveness, the identification of cervical abnormalities, rather than long-term follow-up of outcome, that is reductions in incidence and mortality from cervical cancer.

Data sources

Cost-effectiveness measures require data on natural history of cervical cancer, the overall effectiveness of the policy or intervention, survival rates associated with cancer, test characteristics, and quality of life.

Data sources could include: randomized trials, observational studies, meta-analyses; other published literature, expert opinion and health systems statistics. However, as implied above, surrogate measures of efficacy may have to be used to support assumptions relating to the likely effectiveness of new technology. To the extent that these assumptions are uncertain, the result of the modelling will also be uncertain, even if sensitivity analyses are performed to attempt to encompass the extent of uncertainty.

Sources of cost data could include the costs of:

- Training of staff
- The screening test
- Administration of the screening test
- Laboratory procedures
- Reporting and referral of women with abnormalities
- Diagnostic tests
- Treatment of precursors
- Treatment of clinically invasive cancer
- Patient time for all aspects of screening
- Transportation of specimens
- Programme organization

These data must be collected locally or estimated according to local conditions. The eventual judgement as to whether a particular strategy is cost-effective or not will depend on local health circumstances.

Cost-effectiveness studies

All models consistently support the messages that organized screening is more cost-effective than opportunistic screening and that the most pressing question in all settings is how to reach the highest proportion of women at greatest risk for cervical cancer. Increasing coverage is always more cost-effective than using resources in any other area of the programme. In the United Kingdom, altering the payment system in 1990 as an incentive to primary-care physicians to maximize coverage rather than as a fee for services greatly facilitated the needed increase in coverage, which rose from around 40% of women in 1989 to over 80% of women by 1993 (NHS 2003a, b; Patnick, 2000).

Goldie *et al.* (2001) evaluated the cost-effectiveness of visual inspection, cytology and HPV testing, largely using surrogate measures of efficacy, within a developing country situation. A very important determinant of cost-effectiveness in this setting was the

requirement for three visits for women with abnormal cytological results, whereas for screening by visual inspection a one-visit strategy was modelled, and for HPV testing two visits. A substantial loss when women are required to return after the initial screening visit is observed in many developing countries, and this has an important effect on the cost-effectiveness of cytology, and to a lesser extent of HPV testing. When the authors modelled the effect of a limited number of tests in a lifetime, three tests at five-year intervals from the age of 35 years proved to be more cost-effective than a 10-year schedule commencing at the same age. It was concluded that cervical cancer screening strategies that incorporate DVI or HPV DNA testing and eliminate colposcopy may offer attractive alternatives to cytology-based screening programs in low-resource settings.

Goldie *et al.* (2004) reported the results of an analysis comparing the cost-effectiveness of HPV testing with that of conventional cytology in women aged 30 years or more. This was set in the US context of annual conventional cytological testing, which was compared with three-year screening using liquid-based cytology and three-year screening using HPV testing. [Although the latter strategy proved to be more cost-effective in the analysis, it is unclear what the results would have been if three-yearly conventional cytology had been incorporated in the analysis.] Goldie *et al.* (2004) considered that for women aged 30 years and more, a strategy of screening every two or three years with either HPV DNA testing in combination with cytology for primary screening or cytology with reflex HPV DNA testing for equivocal results will provide a greater reduction in cancer and be less costly than annual conventional cytology.